Weight space structure and analysis using a finite replica number in the Ising perceptron

Tomoyuki Obuchi†§ and Yoshiyuki Kabashima‡

†Department of Physics, Tokyo Institute of Technology,
Tokyo 152-8551, Japan
‡Department of Computational Intelligence and Systems Science,
Tokyo Institute of Technology, Yokohama 226-8502, Japan

Abstract. The weight space of the Ising perceptron in which a set of random patterns is stored is examined using the generating function of the partition function $\phi(n) = (1/N) \log[Z^n]$ as the dimension of the weight vector N tends to infinity, where Z is the partition function and $[\cdots]$ represents the configurational average. We utilize $\phi(n)$ for two purposes, depending on the value of the ratio $\alpha = M/N$, where M is the number of random patterns. For $\alpha < \alpha_s = 0.833\ldots$, we employ $\phi(n)$, in conjunction with Parisi's one-step replica symmetry breaking scheme in the limit of $n \to 0$, to evaluate the complexity that characterizes the number of disjoint clusters of weights that are compatible with a given set of random patterns, which indicates that, in typical cases, the weight space is equally dominated by a single large cluster of exponentially many weights and exponentially many small clusters of a single weight. For $\alpha > \alpha_s$, on the other hand, $\phi(n)$ is used to assess the rate function of a small probability that a given set of random patterns is atypically separable by the Ising perceptrons. We show that the analyticity of the rate function changes at $\alpha = \alpha_{GD} = 1.245\ldots$, which implies that the dominant configuration of the atypically separable patterns exhibits a phase transition at this critical ratio. Extensive numerical experiments are conducted to support the theoretical predictions.

§ obuchi@stat.phys.titech.ac.jp

1. Introduction

The generating function (density) with respect to the partition function Z:

$$\phi(n) = \frac{1}{N} \log[Z^n] \quad (n \in \mathbb{R}), \tag{1}$$

plays a key role in research on disordered systems, where N denotes the size of the objective system and $[\cdots]$ denotes the average over the quenched randomness. Assessing $\phi(n)$ for $\forall n \in \mathbb{R}$ exactly is, in general, difficult, whereas the analytical evaluation for $n = 1, 2, \ldots \in \mathbb{N}$, in conjunction with the use of the saddle point method as $N \to \infty$, is possible for a class of systems. This indicates that (1) can be practically evaluated by analytically continuing the expressions of $\phi(n)$ evaluated for $n \in \mathbb{N}$ to $n \in \mathbb{R}$, which is often referred to as the *replica method.*

In most models of statistical mechanics of disordered systems, the probability that free energy density, $-(1/N) \log Z$, will take a certain value f, $P(f)$, can be expressed in large deviation statistics as

$$P(f) \sim \exp\{NR(f)\}, \tag{2}$$

where $R(f) \leq 0$ is often referred to as the *rate function.* One of recent progresses of the replica theory is the formation of a link between $R(f)$ and $\phi(n)$ [1, 2, 3]. When $R(f)$ is a convex upward function, it can be assessed from $\phi(n)$ being parameterized by $n \in \mathbb{R}$ as

$$f(n) = -\frac{\partial \phi(n)}{\partial n}, \quad R(f(n)) = \phi(n) - n\frac{\partial \phi(n)}{\partial n}. \tag{3}$$

This indicates that the *typical* value of f, which is characterized by the condition $R(f) = (1/N) \log P(f) = 0$, can be evaluated as

$$f^* = -\lim_{n \to 0} \frac{\partial \phi(n)}{\partial n}, \tag{4}$$

which is sometimes referred to as a replica trick formula. Equation (3) indicates that $n = 1, 2, \ldots$ corresponds to *atypical* samples of $R(f) < 0$ representing a small probability. This means that the replica trick can be regarded as a formula that infers the behavior of typical samples by extrapolating the behavior for atypical samples.

Another recent advance in the replica theory is the association between the complex structure of phase space and a formalism of one-step replica symmetry breaking (1RSB) [4, 5, 6]. In a number of systems that are subject to disordered interactions, the phase space is considered to be divisible into exponentially many disjoint sets as $N \to \infty$. Each of the disjoint sets is sometimes referred to as a pure state. Let us assume that the number of pure states specified by the free energy (density) value f, $\mathcal{N}(f)$, is scaled as

$$\mathcal{N}(f) \sim \exp\{N\Sigma(f)\}, \tag{5}$$

where the exponent $\Sigma(f) \geq 0$ is referred to as the *complexity.* Saddle point evaluation of $\sum_\gamma \exp\{-Nxf_\gamma\}$, where $x \in \mathbb{R}$ is a certain control parameter and γ and f_γ are indices

of a pure state and its free energy, respectively, indicates that $\Sigma(f)$ can be evaluated using another generating function $g(x) = (1/N) \log \left(\sum_\gamma \exp\{-Nx f_\gamma\} \right)$ as

$$f(x) = -\frac{\partial g(x)}{\partial x}, \quad \Sigma(f(x)) = g(x) - x \frac{\partial g(x)}{\partial x}, \tag{6}$$

which is parameterized by x as long as $\Sigma(f)$ is convex upward. This formalism is defined for each sample of quenched randomness and, therefore, has nothing to do with $\phi(n)$. However, recent studies have revealed that typical $g(x)$ (more precisely, $(1/x)g(x)$) over the quenched randomness can also be assessed from $\phi(n)$ by evaluation of (4) under Parisi's 1RSB ansatz, handling the 1RSB parameter x as a control parameter.

The concepts of the two exponents $R(f)$ and $\Sigma(f)$ are different in that $R(f) \leq 0$ represents a small probability of atypical samples, whereas $\Sigma(f) \geq 0$ represents a large number of pure states that occur for typical samples. However, the formal similarity of (3) and (6) indicates that there might be a relationship between these two exponents. In fact, when the 1RSB solution of $\phi(n)$, $\phi_{1RSB}(n)$, is assessed using the replica symmetric (RS) solution $\phi_{RS}(n)$ as $\phi_{1RSB}(n) = \mathrm{Extr}_x\{(n/x)\phi_{RS}(x)\}$, where $\mathrm{Extr}_x\{\cdots\}$ denotes the operation of extremization with respect to x, the functional forms of $R(f)$ and $\Sigma(f)$ are in agreement [2, 3]. In addition, the model class for which this property holds is rather wide, and includes random energy models [7, 8] and p-body spin glass models without external fields [2]. This naturally motivates us to further explore more general relationships among $R(f)$, $\Sigma(f)$, and $\phi(n)$, including cases for which the formal accordance of functional forms between $R(f)$ and $\Sigma(f)$ does not hold.

As a concrete effort for the exploration, we herein consider Ising perceptrons that store random input-output patterns. There are two reasons for considering this system. First, the Ising perceptrons can be macroscopically characterized by a few sets of order parameters and are much easier to handle than systems of sparse couplings [9, 10, 11, 12, 13], for which several numerical calculations are required. Despite the simplicity, this model still could exhibit rich behavior in the phase space involving nontrivial RSB phenomena [14, 15], which is highly suitable for our purpose. The second reason is that the meaning of complexity for the perceptrons of finite size is rather clear. For the Ising perceptrons, a pure state at zero temperature can be identified with a stable cluster, the definition of which will be given in section 5, with respect to single spin flips [16, 17, 18]. For samples of small systems, the size of the clusters can be numerically evaluated by exhaustive enumeration without any ambiguity. This property is extremely useful for justifying theoretical predictions through numerical experiments.

The remainder of the present paper is organized as follows. In the next section, we introduce the model that considered herein. In section 3, we provide a formalism that assesses the complexity and rate function based on the 1RSB evaluation of the generating function, for the Ising perceptrons. In the formalism, the complexity and rate function are defined not for the free energy f but for the entropy s, because the analysis is carried out for the micro-canonical ensemble of Ising weights that are perfectly compatible with a given set of random patterns. In section 4, we analyze the behavior of the weight space of the Ising perceptron using this formalism. It is found that for

$\alpha = M/N < \alpha_s = 0.833\ldots$, where M is the number of random patterns, the typical phase space of the Ising perceptron is characterized by a *convex downward* complexity being equally dominated by a single large cluster of exponentially many weights and exponentially many small clusters of a single weight. For $\alpha > \alpha_s$, on the other hand, the rate function becomes relevant for the analysis because random patterns that are perfectly separable by the Ising weights are generated only atypically in this region. It is also found that a certain transition of the rate function occurs at another critical ratio $\alpha_{\text{GD}} = 1.245\ldots$. These predictions are validated by comparison with the results of extensive numerical experiments in section 5. The final section is devoted to a summary.

2. Model definition

A simple perceptron is a map from $\mathbb{R}^N$ to $\{+1, -1\}$ defined as

$$y = \begin{cases} +1, & \boldsymbol{S} \cdot \boldsymbol{x}/\sqrt{N} > 0, \\ -1, & \boldsymbol{S} \cdot \boldsymbol{x}/\sqrt{N} < 0, \end{cases} \tag{7}$$

where $\boldsymbol{x} \in \mathbb{R}^N$ is the input pattern and $y \in \{+1, -1\}$ is the output label. The vector $\boldsymbol{S}$ denotes the adjustable synaptic weight. We hereinafter focus on the case of Ising weight $S_i \in \{+1, -1\}$. In a general scenario, the perceptron stores a given set of M labeled patterns

$$D^M = \{(\boldsymbol{x}_1, y_1), \cdots, (\boldsymbol{x}_M, y_M)\}, \tag{8}$$

by adjusting the weight $\boldsymbol{S}$ so as to completely reproduce the given label y_μ for the input $\boldsymbol{x}_\mu$ for $\mu = 1, 2, \ldots, M$.

In the following, we consider the situation in which the patterns are independently and identically distributed samples from

$$P(\boldsymbol{x}) = \left(\frac{1}{\sqrt{2\pi}}\right)^N \exp\left(-\frac{\boldsymbol{x}^2}{2}\right), \tag{9}$$

$$P(y) = \frac{1}{2}\left(\delta(y-1) + \delta(y+1)\right). \tag{10}$$

The question we address herein is how the space of the weights that store D^M is characterized macroscopically when pattern ratio $\alpha = M/N \sim O(1)$ is fixed as M and N tend to infinity.

3. Formalism

3.1. RS and 1RSB solutions of the generating function

As bases of our analysis, we first provide expressions of RS and 1RSB solutions of the generating function. Since these solutions have been derived numerous times in earlier studies [14, 15], we present only a sketch of the derivation in the main text, and details are shown in Appendix A. For readers who are not familiar with the replica method, we refer to [19, 20].

We first define the Boltzmann factor $\eta(\boldsymbol{S}|D^M)$ of the present system as it takes 1 if the weight $\boldsymbol{S}$ is compatible with D^M and 0 otherwise‖. The explicit form is expressed as

$$\eta(\boldsymbol{S}|D^M) = \prod_{\mu=1}^{M} \Theta\left(-y_\mu \frac{\boldsymbol{S} \cdot \boldsymbol{x}_\mu}{\sqrt{N}}\right), \tag{11}$$

where $\Theta(u) = 1$ for $u > 0$ and $\Theta(u) = 0$, otherwise. The partition function $Z(D^M) \equiv \sum_{\boldsymbol{S}} \eta(\boldsymbol{S}|D^M)$ is equal to the number of weights that are perfectly compatible with D^M in this situation, and varies randomly depending on the quenched randomness D^M. This naturally leads us to evaluate the generating function $\phi(n) = (1/N) \log \left[Z^n(D^M)\right]_{D^M}$ using the replica method, where $[\cdots]_{D^M}$ represents the operation of averaging with respect to D^M. For $n = 1, 2, \ldots \in \mathbb{N}$, this yields the following expression:

$$\phi(n) = \operatorname*{Extr}_{q^{ab}, \widehat{q}^{ab}} \left\{ -\sum_{a<b} \widehat{q}^{ab} q^{ab} + \log \left(\sum_{S^1, S^2, \ldots, S^n} e^{\sum_{a<b} \widehat{q}^{ab} S^a S^b} \right) \right.$$
$$\left. + \alpha \log \left[\prod_{a=1}^{n} \Theta(u^a) \right]_{\boldsymbol{u}} \right\}, \tag{12}$$

where $[\cdots]_{\boldsymbol{u}}$ represents averaging with respect to multivariate Gaussian random variables $u^1, u^2, \ldots, u^n$, the first and second moments of which are specified as $[u^a]_{\boldsymbol{u}} = 0$ and $\left[u^a u^b\right]_{\boldsymbol{u}} = \delta_{ab} + (1 - \delta_{ab}) q^{ab}$ $(a, b = 1, 2, \ldots, n)$, respectively.

Analytical continuation from $n \in \mathbb{N}$ to $n \in \mathbb{R}$ is performed by imposing a certain permutation symmetry on the extremum point of the right-hand side of (12). We find several solutions in the RS and 1RSB levels.

3.1.1. RS solutions Constraints $q^{ab} = q$ and $\widehat{q}^{ab} = \widehat{q}$ characterize the RS solutions. Solving the extremization problem of (12) analytically and numerically under these constraints yields the following two solutions:
RS1: $0 < q < 1$ and $\widehat{q} < +\infty$.

$$\phi_{\mathrm{RS1}}(n) = -\frac{n(n-1)}{2} q\widehat{q} - \frac{1}{2} n\widehat{q} + \log \left(\int Dz \left(2 \cosh \left(\sqrt{\widehat{q}} z \right) \right)^n \right)$$
$$+ \alpha \log \left(\int Dz E^n \left(\sqrt{\frac{q}{1-q}} z \right) \right), \tag{13}$$

where $Dz = \exp\left(-z^2/2\right)/\sqrt{2\pi}$ represents the Gaussian measure and $E(u) = \int_u^{+\infty} Dz$.
RS2: $q = 1$ and $\widehat{q} = +\infty$.

$$\phi_{\mathrm{RS2}}(n) = (1 - \alpha) \log 2. \tag{14}$$

‖ This is equivalent to the zero temperature limit $\beta \to \infty$ of the Boltzmann factor $e^{-\beta H(\boldsymbol{S}|D^M)}$ where the Hamiltonian $H(\boldsymbol{S}|D^M)$ is given by $\sum_{\mu=1}^{M} \Theta\left(-y_\mu \frac{\boldsymbol{S} \cdot \boldsymbol{x}_\mu}{\sqrt{N}}\right)$, which is equal to the number of patterns that are incompatible with the weight $\boldsymbol{S}$.

3.1.2. 1RSB solutions In 1RSB solutions, replica indices are divided into n/m groups of identical size m. Constraints for characterizing the 1RSB solutions are expressed as

$$q^{ab} = \begin{cases} q_1 & \text{if } a \text{ and } b \text{ belong to the same group,} \\ q_0 & \text{otherwise,} \end{cases} \tag{15}$$

and are similarly expressed for $\widehat{q}^{ab}$. Three solutions are found under these constraints:
1RSB1: $(q_1, q_0) = (1, q)$ and $(\widehat{q}_1, \widehat{q}_0) = (+\infty, \widehat{q})$, where q and $\widehat{q}$ take the same values as those for $\phi_{\mathrm{RS1}}(n)$.

$$\phi_{\mathrm{1RSB1}}(n, m) = \phi_{\mathrm{RS1}}\left(\frac{n}{m}\right). \tag{16}$$

1RSB2: $(q_1, q_0) = (q, q)$ and $(\widehat{q}_1, \widehat{q}_0) = (\widehat{q}, \widehat{q})$, where q and $\widehat{q}$ take the same values as those for $\phi_{\mathrm{RS1}}(n)$.

$$\phi_{\mathrm{1RSB2}}(n, m) = \phi_{\mathrm{RS1}}(n). \tag{17}$$

1RSB3: $(q_1, q_0) = (1, 1)$ and $(\widehat{q}_1, \widehat{q}_0) = (+\infty, +\infty)$.

$$\phi_{\mathrm{1RSB3}}(n, m) = \phi_{\mathrm{RS2}}(n) = (1 - \alpha) \log 2. \tag{18}$$

In usual analyses, Parisi's 1RSB parameter m is determined by the extremum condition in evaluating $\phi(n) = \mathrm{Extr}_m \{\phi_{\mathrm{1RSB*}}(n, m)\}$, where $* = 1, 2$ and 3. In addition, there might be no need to classify **1RSB2** and **1RSB3** as 1RSB solutions because **1RSB2** and **1RSB3** are completely reduced to **RS1** and **RS2**, respectively. However, handling these three solutions as 1RSB solutions, leaving the m-dependence of $\phi_{\mathrm{1RSB}}(n, m)$ explicitly, is crucial for the current purpose of relating the concepts of $\phi(n)$, $\Sigma(s)$, and $R(s)$ based on physical considerations presented in the following subsection.

3.2. 1RSB solution as the generating function of complexity and rate function

Let us present the 1RSB solutions through a physical inference based on arguments presented in earlier studies [5, 6, 12, 16, 17]. Here, we assume a situation in which the weight space is divided into exponentially many pure states for a given sample of D^M.

We introduce an indicator function $\delta_\gamma(\boldsymbol{S})$, which is defined as $\delta_\gamma(\boldsymbol{S}) = 1$, if $\boldsymbol{S}$ belongs to the pure state γ and 0, otherwise, to express the number of weights included in γ as

$$Z_\gamma = \sum_{\boldsymbol{S}} \prod_{\mu=1}^{M} \Theta\left(-y_\mu \frac{\boldsymbol{S} \cdot \boldsymbol{x}_\mu}{\sqrt{N}}\right) \delta_\gamma(\boldsymbol{S}). \tag{19}$$

Let us assume that Z_γ typically scales as $Z_\gamma \sim \exp(Ns)$, where $s \sim O(1)$ has the physical meaning of entropy (density), and the number of pure states corresponding to the value of the entropy s increases as $\mathcal{N}(s) \sim \exp(N\Sigma(s))$, where $\Sigma(s) \geq 0$ is the complexity for the entropy s. This assumption, in conjunction with the saddle point assessment, provides us with a generating function of $\Sigma(s)$, as follows:

$$g(x|D^M) = \frac{1}{N} \log\left(\sum_\gamma Z_\gamma^x\right) \sim \frac{1}{N} \log\left(\int ds e^{N(xs + \Sigma(s))}\right)$$

$$= \max_s \left\{ xs + \Sigma(s) \right\}. \tag{20}$$

This relationship indicates that when $\Sigma(s)$ is a convex upward function, it can be assessed from $g(x|D^M)$ as

$$s(x) = \frac{\partial g(x|D^M)}{\partial x}, \quad \Sigma(s(x)) = g(x|D^M) - x\frac{\partial g(x|D^M)}{\partial x}, \tag{21}$$

being parameterized by x. Here, $g(x|D^M)$ is defined for each sample of D^M. However, the self-averaging property is assumed to hold in the current system, which means that $g(x|D^M)$ for typical samples converges to its average $g(x) = \left[g(x|D^M)\right]_{D^M}$ in a large system limit of $N, M \to \infty$ while maintaining $\alpha = M/N \sim O(1)$.

The replica method can be used to assess $g(x)$. For this, we consider the following identity:

$$g(x) = \frac{1}{N}\left[\log\left(\sum_\gamma Z_\gamma^x\right)\right]_{D^M} = \lim_{y\to 0}\frac{\partial}{\partial y}\left(\frac{1}{N}\log\left[\left(\sum_\gamma Z_\gamma^x\right)^y\right]_{D^M}\right). \tag{22}$$

Although exact evaluation of the right-hand side of (22) is difficult, for $x, y \in \mathbb{N}$, the equation (19) and the formula of series expansion provide the following expression:

$$\left[\left(\sum_\gamma Z_\gamma^x\right)^y\right]_{D^M}$$
$$= \sum_{\{\gamma^\sigma\}}\sum_{\{\boldsymbol{S}^{\sigma a}\}}\left[\prod_{\mu=1}^M\prod_{\sigma=1}^y\prod_{a=1}^x\Theta\left(-y_\mu\frac{\boldsymbol{S}^{\sigma a}\cdot\boldsymbol{x}_\mu}{\sqrt{N}}\right)\right]_{D^M}\prod_{\sigma=1}^y\prod_{a=1}^x\delta_{\gamma^\sigma}(\boldsymbol{S}^{\sigma a}), \tag{23}$$

which can be evaluated by the saddle point method in the large system limit.

The following observations are noteworthy in the evaluation.

- The summation is taken over all possible configurations of xy replica weights.
- However, the factor of $\prod_{\sigma=1}^y\prod_{a=1}^x\delta_{\gamma^\sigma}(\boldsymbol{S}^{\sigma a})$ allows only contributions from configurations in which xy replica weights are equally assigned to y pure states by x.

These observations are nothing more than the physical meaning of the 1RSB ansatz in assessing $\left[Z^n(D^M)\right]_{D^M}$ with substitution of $n = xy$ and $m = x$ (figure 1). Accepting this interpretation yields the following expression:

$$\frac{1}{N}\log\left[\left(\sum_\gamma Z_\gamma^x\right)^y\right]_{D^M} = \phi_{1\mathrm{RSB}}(xy, x), \tag{24}$$

where $\phi_{1\mathrm{RSB}}(n, m)$ is the 1RSB solution considered in the previous section, and its concrete functional form should be chosen appropriately from among **1RSB1**, **1RSB2**, and **1RSB3** for a given pair of α and n.

Inserting (24) into (22) yields $g(x) = x(\partial/\partial n)\phi_{1\mathrm{RSB}}(n, x)|_{n=0}$, which directly yields the following formula relating typical complexity to $\phi_{1\mathrm{RSB}}(n, m)$:

$$\begin{cases} s(x) = (\partial/\partial x)\left(x(\partial/\partial n)\phi_{1\mathrm{RSB}}(n, x)|_{n=0}\right), \\ \Sigma(s(x)) = -x^2(\partial^2/\partial x\partial n)\phi_{1\mathrm{RSB}}(n, x)|_{n=0}. \end{cases} \tag{25}$$
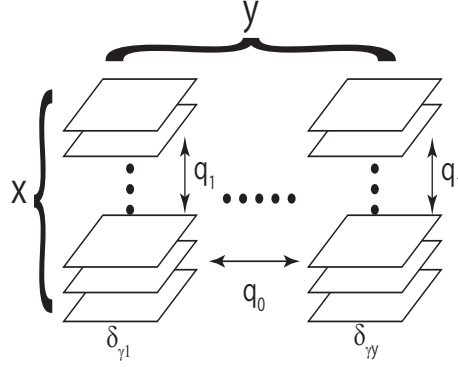
Figure 1. Schematic diagram of the 1RSB structure of the factor $\prod_{\sigma=1}^{y} \prod_{a=1}^{x} \delta_{\gamma^{\sigma}}(S^{\sigma a})$.

On the other hand, an identity with respect to the indicator function $\sum_{\gamma} \delta_{\gamma}(\boldsymbol{S}) = 1$ for $\forall \boldsymbol{S}$ guarantees $\sum_{\gamma} Z_{\gamma} = Z(D^M)$, indicating that $\phi(n) = \phi_{1RSB}(n, x)|_{x=1}$ holds in general. This means that the rate function can be assessed from $\phi_{1RSB}(n, m)$ as follows:

$$\begin{cases} s_{tot}(n) = (\partial/\partial n)\phi_{1RSB}(n, x)|_{x=1}, \\ R(s_{tot}(n)) = -n^2(\partial/\partial n)\left(n^{-1}\phi_{1RSB}(n, x)|_{x=1}\right), \end{cases} \quad (26)$$

where we define the total entropy $s_{tot} \equiv \lim_{N\to\infty}(1/N)\log Z = \max_s\{s + \Sigma(s)\}$ which corresponds to the total number of weights that are compatible with D^M. In (25), the parameter x can vary only in such a range that both $s(x) \geq 0$ and $\Sigma(s(x)) \geq 0$ hold. Similarly, the conditions $s_{tot}(n) \geq 0$ and $R(s_{tot}(n)) \leq 0$ restrict the range of n in (26). These constitute the main result of the present paper.

Here, three issues are noteworthy. First, for a class of disordered systems, including random energy models and p-body spin glass models without external fields, two equalities $\phi_{1RSB}(n, m) = (n/m)\phi_{RS}(m)$ and $\phi_{1RSB}(n, m = 1) = \phi_{RS}(n)$, hold in assessing the complexity and rate function, respectively, where $\phi_{RS}(n)$ is an identical RS solution of the generating function $\phi(n)$. Inserting these functions into (25) and (26) offers an identical functional form for both the complexity and the rate function, while their domains of definition are disjointed, except for a point of the typical value of free energy f^* (or entropy s^*). The current system, however, does not possess this property because $\phi_{1RSB}(n, m) = (n/m)\phi_{RS}(m)$ does not hold for **1RSB1**, **1RSB2**, or **1RSB3** while $\phi_{1RSB}(n, m = 1) = \phi_{RS}(n)$ is always satisfied. Second, (25) and (26) are valid only when $\phi_{1RSB}(n, m)$ are stable against any perturbation for a further RSB. Fortunately, in the present problem, a stable solution against any known RSB instabilities can be constructed for $\forall \alpha > 0$ and $\forall n > 0$. This implies that, in the present analysis, there is no need to consider further RSB. Finally, however, we have to keep in mind that (25) and (26) depend on the assumptions that correct $\Sigma(s)$ and $R(s)$ are convex upward functions, respectively. When the convex upward property does not hold, the estimates of (25) and (26) represent not the correct solution, but rather its convex hull. The following analytical and experimental assessment indicates that this is the case for $\Sigma(s)$

of sufficiently low α and $R(s)$ of sufficiently high α.

4. Theoretical predictions

We are now ready to use the formalism developed above to analyze the behavior of the weight space of the Ising perceptron.

4.1. Complexity for $\alpha < \alpha_s = 0.833\ldots$

In order to perform the analysis, it is necessary to select a certain solution (functional form) from among the three candidates of **1RSB1**, **1RSB2**, and **1RSB3**. Analyticity and physical plausibility are two guidelines for this task.

   The replica method is a scheme to infer the properties for real replica numbers $n \in \mathbb{R}$ by analytical continuation from those for natural numbers $n = 1, 2, \ldots \in \mathbb{N}$. This indicates that, for examining typical $(n \to 0)$ behavior, it is plausible to select the solution of $\phi(n)$ that is dominant around $n \geq 1$, because unity is the natural number that is closest to zero. For $\alpha < \alpha_s = 0.833\ldots$, this solution is $\phi_{\mathrm{RS1}}(n)$. In addition, the relevant $\phi_{\mathrm{1RSB}}(n, m)$ must agree with this solution at $m = 1$. These considerations offer two candidates of $g(x)$ as

$$g_{\mathrm{1RSB1}}(x) = x \frac{\partial}{\partial n} \phi_{\mathrm{1RSB1}}(n, x)|_{n=0} = \phi'_{\mathrm{RS1}}(0), \tag{27}$$

and

$$g_{\mathrm{1RSB2}}(x) = x \frac{\partial}{\partial n} \phi_{\mathrm{1RSB2}}(n, x)|_{n=0} = x \phi'_{\mathrm{RS1}}(0). \tag{28}$$

   We combine these solutions to construct an entire functional form of $g(x)$ based on physical considerations. For $x \gg 1$, $g(x)$ should vary approximately linearly with respect to x, because a single pure state of the largest entropy typically dominates $\sum_\gamma Z_\gamma^x$. In addition, $s(x) = (\partial/\partial x)g(x)$ for $x \sim 0$ should be smaller than that for $x \gg 1$ because $s(x)$ should increase monotonically with respect to x. Furthermore, $g(x)$ must be a continuous function. These considerations reasonably yield an entire functional form of $g(x)$ as

$$g(x) = \begin{cases} \phi'_{\mathrm{RS1}}(0), & x \leq 1, \\ x \phi'_{\mathrm{RS1}}(0), & x > 1, \end{cases} \tag{29}$$

which yields the complexity as

$$\Sigma(s) = \begin{cases} \phi'_{\mathrm{RS1}}(0) - s, & 0 \leq s \leq \phi'_{\mathrm{RS1}}(0), \\ -\infty, & \text{otherwise.} \end{cases} \tag{30}$$

   The piecewise linear profile of (29) is somewhat extraordinary. This is thought to be because the correct complexity is not convex upward in this system. When $\Sigma(s)$ is convex upward, the current formalism using the saddle-point method defines a one-to-one map between $g(x)$ and $\Sigma(s)$. However, if $\Sigma(s)$ is not convex upward, the functional profile of a region in which the correct complexity is convex downward is lost and only

the convex hull is obtained by the transformation from $g(x)$, as shown in figure 2. The piecewise liner profile of $g(x)$ presumably signals that this actually occurs in the current problem. Similar behavior of the complexity could also be observed in a certain type of random energy models [21].

The physical implication of (30), the profile of which is obtained by connecting two points $(s, \Sigma) = (0, \phi'_{\text{RS1}}(0))$ and $(\phi'_{\text{RS1}}(0), 0)$ with a straight line having a slope of $-x = -1$, is that the weight space is equally dominated by exponentially many clusters of vanishing entropy and a subexponential number of large clusters composed of exponentially many weights. The existence of large clusters may accord with an earlier study which reported that local search heuristics of a certain type manage to find a compatible weight efficiently up to a considerably large value of α near to the capacity α_s [22]. On the other hand, the coexisting exponentially many small clusters may be a major origin of a known difficulty in finding compatible weights by Monte Carlo sampling schemes [23, 24].
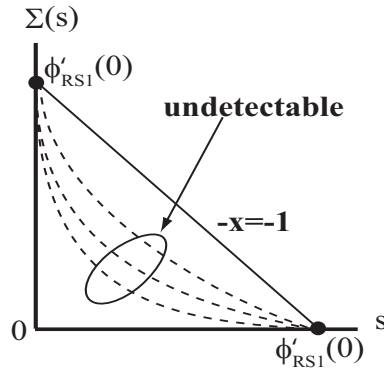

Figure 2. Schematic profile of complexity (30). The characteristic exponent of the size distribution of pure states cannot be correctly assessed in the current formalism if it is a convex downward function (dashed curves). In such cases, the complexity $\Sigma(s)$ assessed from $g(x)$ (solid line) is the convex hull (black circle) of the correct exponent.

4.2. Rate function for $\alpha > \alpha_{\text{s}} = 0.833 \ldots$ and a transition at $\alpha_{\text{GD}} = 1.245 \ldots$

For $\alpha_{\text{s}} < \alpha$, (30) becomes negative, which implies that there exist no compatible weights for typical samples of D^M. In such cases, the rate function $R(s)$, which characterizes a small probability that atypical samples that are compatible with the Ising perceptrons are generated, becomes relevant in the current analysis. Therefore, we focus on the assessment of this exponent for this region.

For $\alpha_{\text{s}} < \alpha < \alpha_{\text{GD}} = 1.245 \ldots$, $\phi_{\text{RS1}}(n)$ dominates the generating function $\phi(n)$ in the vicinity of $n \geq 1$ as for $\alpha < \alpha_{\text{s}}$. This means that $\phi_{\text{1RSB1}}(n, m = 1) = \phi_{\text{1RSB2}}(n, m = 1) = \phi_{\text{RS1}}(n)$ should be used to assess $R(s)$ of relatively frequent events that correspond to $0 < n < 1$. However, this function is minimized to a negative value at a certain point at which $0 < n_{\text{s}}(\alpha) < 1$, which implies that assessment by naïvely using

$\phi_{\mathrm{RS1}}(n)$ for $n < n_{\mathrm{s}}(\alpha)$ leads to incorrect results, which yield a negative total entropy $s_{\mathrm{tot}}(n) = (\partial/\partial n)\phi_{\mathrm{RS}}(n) < 0$. In order to avoid this inconsistency, we fix the value of $\phi(n)$ to $\phi_{\mathrm{RS1}}(n_{\mathrm{s}}(\alpha))$, which is reduced to the conventional construction of a frozen RSB solution. In particular, this yields an assessment of

$$R(0) = \phi_{\mathrm{RS1}}(n_{\mathrm{s}}(\alpha)) = \min_{n}\{\phi_{\mathrm{RS1}}(n)\}, \tag{31}$$

which has the physical meaning of a characteristic exponent of a small probability that a given sample set D^M is separable by certain Ising perceptrons. For $\alpha \geq \alpha_{\mathrm{GD}} = 1.245\ldots$, on the other hand, the dominant solution of $\phi(n)$ in the vicinity of $n \geq 1$ is updated from $\phi_{\mathrm{RS1}}(n)$ to $\phi_{\mathrm{RS2}}(n) = (1 - \alpha)\log 2$, which yields

$$R(0) = \phi_{\mathrm{RS2}}(n) = (1 - \alpha)\log 2. \tag{32}$$

In order to provide a visual representation of the above discussions, we depict the behaviors of $\phi(n)$ in figure 3.
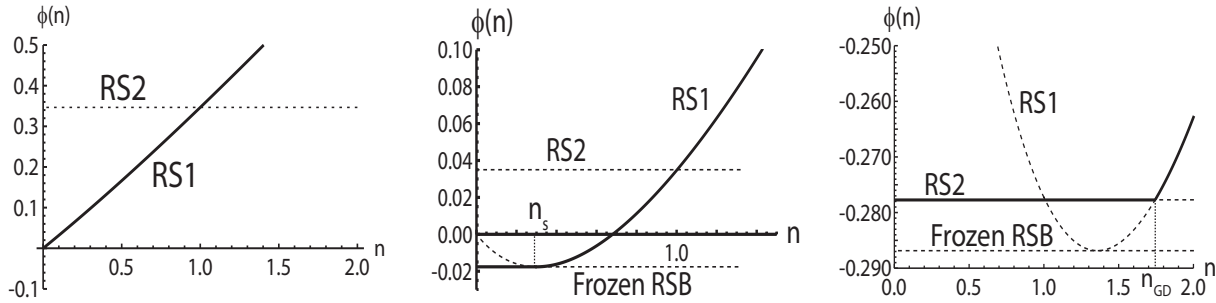

Figure 3. Behavior of $\phi(n)$. The solid lines denote the correct $\phi(n)$, and the dotted lines are the RS and frozen RSB branches. The corresponding values of the parameter α are $0.5, 0.95$, and 1.4, from left to right.

The difference in physical behavior between $\alpha_{\mathrm{s}} < \alpha < \alpha_{\mathrm{GD}}$ and $\alpha > \alpha_{\mathrm{GD}}$ is expected to be as follows. For $\alpha_{\mathrm{s}} < \alpha < \alpha_{\mathrm{GD}}$, the dominant solution around $n > n_{\mathrm{s}}(\alpha)$, $\phi(n) = \phi_{\mathrm{RS1}}(n)$, varies smoothly. This leads to the following behavior of $R(s)$ in the vicinity of $s = 0$:

$$R(s) = R(0) - As^2 + \ldots, \tag{33}$$

where $A > 0$ is a certain constant, which implies that large clusters can appear with a relatively large probability although typical samples of D^M are not separable by the Ising perceptrons. On the other hand, for $\alpha > \alpha_{\mathrm{GD}}$, $\phi(n) = \phi_{\mathrm{RS2}}(n)$ is constant for $n < n_{\mathrm{GD}}(\alpha)$, which is characterized by $\phi_{\mathrm{RS2}}(n_{\mathrm{GD}}(\alpha)) = \phi_{\mathrm{RS1}}(n_{\mathrm{GD}}(\alpha))$ and $n_{\mathrm{GD}}(\alpha) > 1$, and is switched to $\phi(n) = \phi_{\mathrm{RS1}}(n)$ for $n > n_{\mathrm{GD}}(\alpha)$ at $n = n_{\mathrm{GD}}(\alpha)$, which is accompanied by a jump in the first derivative. This indicates that (upward) convexity does not hold for $R(s)$ in the region of $0 < s < (\partial/\partial n)\phi_{\mathrm{RS1}}(n_{\mathrm{GD}}(\alpha))$ as was mentioned for $\Sigma(s)$ in the previous subsection, which implies that the events of $s = 0$ overwhelm those of $s > 0$ in relative probabilities. Therefore, the generation of large clusters should be considerably rare for α of this region.

4.3. Phase diagram on the n-α plane

The above considerations are sufficient to draw a phase diagram on the n-α plane, which is depicted in figure 4.
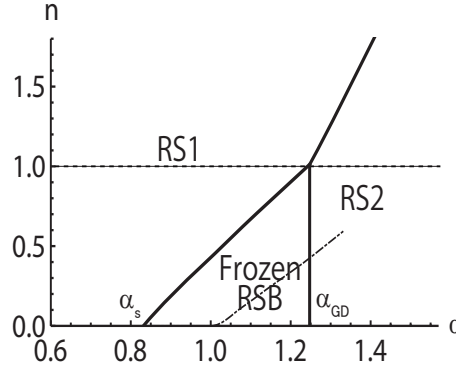

Figure 4. Phase diagram on the n-α plane. Solid lines are phase boundaries, and the dotted line denotes $n = 1$. The dotted-dashed line expresses the AT line for the RS1 solution, but is irrelevant. The AT line vanishes at a certain value of α, because the solution for $0 < q < 1$ vanishes at this point. The RS2 solution involves the AT instability only on the $n = 0$ line, which is presumably of no relevance in the replica analysis.

The value of the tricritical point $\alpha_{\mathrm{GD}} = 1.245\ldots$ is identical to the critical ratio of the perfect learning of the Ising perceptrons in the teacher-student scenario [25, 26]. Formally, this agreement is explained as follows. The dominant solution for $n < 1$ is determined by whether $\phi_{\mathrm{RS1}}(n)$ or $\phi_{\mathrm{RS2}}(n)$ dominates around $n \geq 1$. Since $\phi_{\mathrm{RS1}}(n = 1) = \phi_{\mathrm{RS2}}(n = 1)$ is always guaranteed, the critical condition is given as $(\partial/\partial n)\phi_{\mathrm{RS1}}(n)|_{n=1} = (\partial/\partial n)\phi_{\mathrm{RS2}}(n)|_{n=1} = 0$. On the other hand, $(\partial/\partial n)\phi_{\mathrm{RS1}}(n)|_{n=1}$ generally provides the total entropy after learning in the teacher-student scenario, the target of which can be dealt with as an $(n+1)$-replicated system, in which the teacher is handled as an extra replica. Therefore, the condition of perfect learning, which indicates that the weight of the student agrees perfectly with that of the teacher after learning, is identical to the vanishing entropy condition of the $(n + 1)$-replicated system in the limit $n \to 0$, which agrees with $(\partial/\partial n)\phi_{\mathrm{RS1}}(n)|_{n=1} = 0$, giving the critical value $\alpha_{\mathrm{GD}}(\alpha)$ in the current problem. Although the agreement is justified formally in this manner, its physical implication remains somewhat unclear. The line $n = 1$, which passes through the tricritical point, may have an analogous relation to the concept of Nishimori's line in the theory of spin glasses [19, 27].

Finally, we mention the de Almeida-Thouless (AT) condition in this model [28]. The AT (stability) condition of $\phi_{\mathrm{RS}}(n)$ with the order parameters q and $\widehat{q}$ is expressed as follows:

$$\frac{\alpha}{(1-q)^2} \frac{\int Dz E^n \left(\frac{E''}{E} - \left(\frac{E'}{E}\right)^2\right)^2}{\int Dz E^n} \frac{\int Dz \cosh^{n-4} \sqrt{\widehat{q}}z}{\int Dz \cosh^n \sqrt{\widehat{q}}z} \leq 1. \tag{34}$$

An outline of the derivation is given in [14]. This condition for $\phi_{\mathrm{RS1}}(n)$ is broken in a certain region on the n-α plane, but is irrelevant because the region is always included in $n < n_s(\alpha)$, for which the relevant solution is already switched to that of the frozen RSB. On the other hand, ϕ_{RS2} is stable for $n > 0$ but becomes unstable only on $n = 0$, as reported in [14]. The relevance of this instability for $\alpha \geq \alpha_{\mathrm{GD}}$ may require more a detailed discussion, but we assume herein that this instability can be ignored because only the asymptotic behavior of $\phi(n)$ in the limit $n \to 0$ is relevant in procedures of the replica method.

5. Numerical validation

For validating the theoretical predictions obtained in the previous section, we carried out extensive numerical experiments. In describing the experiments, let us first define the cluster in the present problem. The cluster is a set of spin configurations that are stable with respect to single spin flips [16, 17, 18]. Clusters have the following properties:

- Any configuration belongs to a cluster.
- When a spin configuration "A" can be moved to another configuration "B" by a single spin flip without changing the number of incompatible patterns, "A" and "B" belong to the same cluster.

In the following, we concentrate on vanishing energy clusters, which are composed of weights that are perfectly compatible with D^M.

Before going into details, we elucidate the relation between the cluster and the pure state. Identifying the microscopic description of a pure state is generally a delicate problem, but in the Ising perceptron a pure state can be identified with a cluster, as mentioned in section 1. There is no proof of this statement but it is naturally understood by considering the following aspects of the present problem: The Boltzmann weight of $\eta(\boldsymbol{S}|D^M)$ in (11) becomes completely zero if there is any incompatible pattern. This means that accessing from a cluster to a different cluster by single spin flips is impossible because those clusters are completely separated by states with zero probability $\eta(\boldsymbol{S}|D^M) = 0$. This naturally leads to identifying a cluster with a pure state, because a pure state is a set of configurations which cannot be accessed from other sets by natural dynamics. Several earlier studies support this description [16, 17, 18], and we hereafter admit this assumption.

Now, let us return to the experiments. We denote the size of a cluster as Q and the number of size-Q clusters for a sample D^M as $C(Q|D^M)$. the entropy of a cluster s is considered to be identified by $s = (1/N) \log Q$, and the complexity $\Sigma(s|D^M)$ corresponds to $(1/N) \log C(Q|D^M)$. The clusters can be numerically evaluated, and hence we can construct the 1RSB generating function from the numerical data as

$$\phi_{\mathrm{1RSBnum}}(n = xy, m = x) = \frac{1}{N} \log \left[\left(\sum_{\gamma} Q_{\gamma}^x \right)^y \right], \tag{35}$$

where $[\ldots]$ denotes the sample average operation with respect to D^M. In the typical limit $y \to 0$, this yields the following expression:

$$g_{\text{num}}(x) = \lim_{y \to 0} \frac{\partial}{\partial y} \phi_{\text{1RSBnum}}(xy, x) = \frac{1}{N} \frac{\left[\Theta\left(\sum_\gamma Q_\gamma^x \right) \log \left(\sum_\gamma Q_\gamma^x \right) \right]}{\left[\Theta\left(\sum_\gamma Q_\gamma^x \right) \right]}, \quad (36)$$

where the step function $\Theta(x)$ comes from the differentiation of $\log\left[\left(\sum_\gamma Q_\gamma^x \right)^y \right]$ with respect to y. This means that if there is no cluster for a sample D^M, then the contribution of $\Theta\left(\sum_\gamma Q_\gamma^x \right) \log \left(\sum_\gamma Q_\gamma^x \right)$ vanishes.

In order to examine the consistency with the replica analysis, we assess (36) based on data obtained in extensive numerical experiments. The function $g_{\text{num}}(x)$ is evaluated by the exact enumeration of weights that are compatible with D^M, which are referred to hereinafter as *solutions*. The procedure is summarized as follows:

(i) Generate M examples $D^M = \{(y_1, \boldsymbol{x}_1) \cdots (y_M, \boldsymbol{x}_M)\}$.

(ii) Enumerate all solutions.

(iii) Partition the solutions into clusters, and calculate $\sum_\gamma Q_\gamma^x$ for an appropriate set of x. We actually took 41 points between $x = 0$ and 2.0.

(iv) Repeat the above procedures until sufficient data are obtained and calculate $\left[\Theta\left(\sum_\gamma Q_\gamma^x \right) \log \left(\sum_\gamma Q_\gamma^x \right) \right]$ by taking the sample average.

The resultant plots of $g_{\text{num}}(x)$ for $\alpha = 0.5$ are shown in figure 5. As the system size
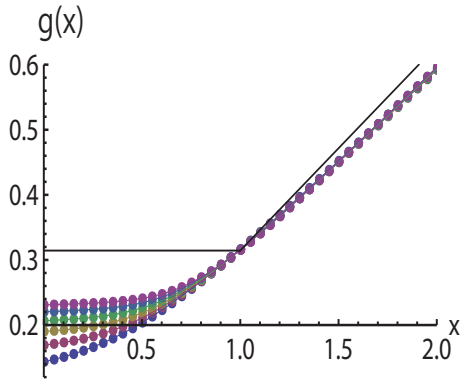

Figure 5. Behavior of $g_{\text{num}}(x)$ for $\alpha = 0.5$. The system sizes are $N = 14, 16, 18, 20, 22,$ and 24, from bottom to top. The solid lines denote $g(x)$ given by (29). The number of samples is $32,000$ for each N. Error bars are smaller than the size of markers. As the system size grows, the profiles of $g_{\text{num}}(x)$ approach the theoretical prediction.
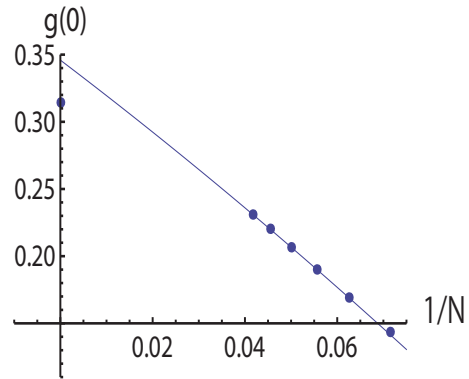
Figure 6. Size dependence of $g_{\text{num}}(0)$. The data are the same as figure 5. The solid line is provided by quadratic fitting to the data. The point at $1/N = 0$ is the theoretical value derived by the replica method. The data tend to reach this theoretical value as the system size grows.

grows, the numerical data for $x \leq 1$ exhibit flatter slopes approaching the theoretical prediction $g(x) = \phi'_{\mathrm{RS1}}(0)$ for $x < 1$. This can also be seen in figure 6 as the systematic approaching of $g_{\mathrm{num}}(0)$ to the theoretical value of $g(0) = 0.314\ldots$ derived from the replica analysis. The difference between the numerical extrapolation and the analytical result at $1/N = 0$ is considered to be the systematic error due to higher order contributions of $1/N$. The profiles of $x > 1$, on the other hand, are approximately straight lines, and the slopes appear gentle than that of the theoretical prediction $\phi'_{\mathrm{RS1}}(0)$. However, the data still slowly move closer to $x\phi'_{\mathrm{RS1}}(0)$ $(x > 1)$ as N becomes larger as a whole, implying consistency with the theoretical prediction.

Complexity $\Sigma(s)$ can also be assessed from the numerical data. One scheme for evaluating $\Sigma(s)$ is to use the relations of (25) with a polynomial interpolation of the numerical data. We determined the order of the polynomial using Akaike's information criteria [29] and eventually selected a 27th degree polynomial, but the obtained results were not so sensitive to details of the choice of the polynomial. The assessed profiles of $\Sigma(s)$ are plotted in figure 7. The curves appear to approach the line predicted in the
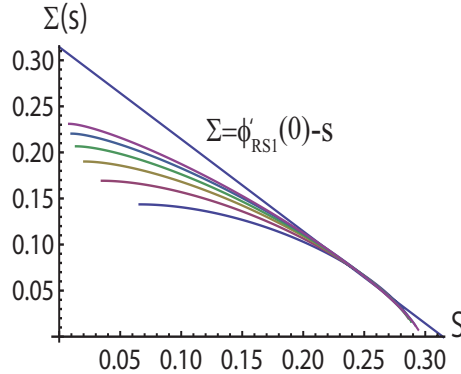

Figure 7. $\Sigma(s(x))$ obtained from $g_{\mathrm{num}}(x)$ using the relation of (25) for $\alpha = 0.5$. The system sizes increase from bottom to top. The solid line denotes the asymptotic line in the thermodynamic limit predicted by the replica analysis.

previous section as N increases, which supports our replica analysis.

However, the complexity curve shown in figure 7 might lose the information about the correct distribution of the clusters, as mentioned in section 4.2. In order to examine this possibility, we directly evaluate the distribution of pure states in a rather naive manner. We refer to the result of this assessment as the *raw complexity*, which is defined as

$$\Sigma_{\mathrm{r}}(s = (1/N)\log Q|D^M) = \frac{1}{N}\Theta\left(C(Q|D^M)\right)\log\left(C(Q|D^M)\right). \qquad (37)$$

Taking the sample average yields the typical profile of $\Sigma_{\mathrm{r}}(s|D^M)$ as $\Sigma_{\mathrm{r}}(s) = [\Sigma_{\mathrm{r}}(s|D^M)]$, the result of which for $\alpha = 0.5$ is shown in figure 8. We took $32,000$ samples in the evaluation for each size and joined the plots to obtain smooth curves. This figure indicates that $\Sigma_{\mathrm{r}}(0)$ approaches the value of the theoretical prediction $\phi'_{\mathrm{RS1}}(0)|_{\alpha=0.5} = 0.314\ldots$ from below as N increases. However, $\Sigma_{\mathrm{r}}(x)$ for $x \geq 0.1$ appears to remain
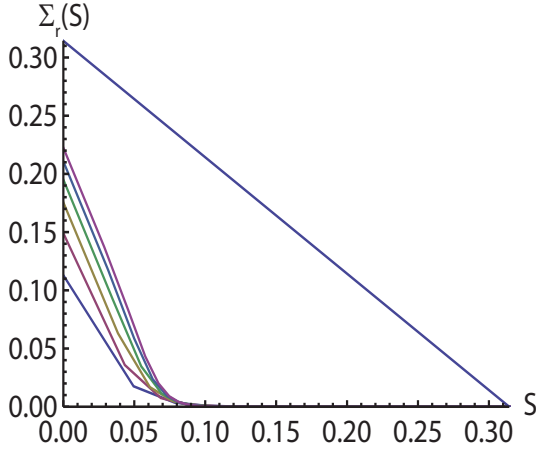
Figure 8. Plot of the raw complexity $\Sigma_r(s)$ for $\alpha = 0.5$. The system size increases from $N = 14$ to 24 in increments of 2, from bottom to top. The solid line is the same as that shown in figure 7.
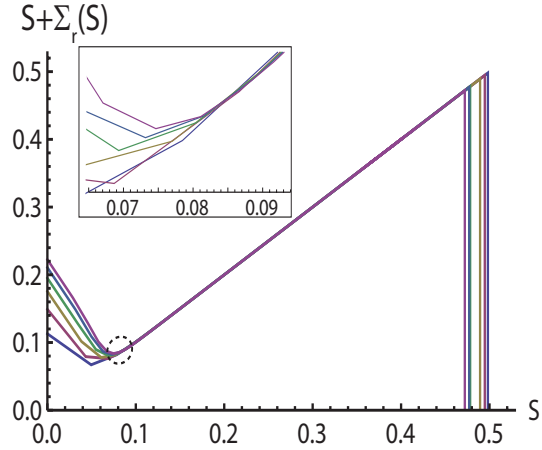
Figure 9. Plot of $s + \Sigma_r(s)$. As the system size increases, the curve appears to converge to a V-shape function, indicating that $\Sigma_r(s)$ is convex downward. The inset shows a close-up of the region enclosed by the dotted ellipse.

approximately constant at zero, indicating that $\Sigma_r(x)$ converges to a convex downward function. We also plot the function $s + \Sigma_r(s)$ in figure 9. This plot shows two peaks and one dip of $s + \Sigma_r(s)$, indicating that $\Sigma_r(s)$ is a convex downward function. The position of the right-hand peak tends to move left to the right terminal point $s = 0.314\ldots$ of the theoretical prediction as the system size increases, while the dip appears to be bounded at the point $x = 0.084$ as shown in the inset. In conclusion, these figures indicate that the exponent that characterizes the size distribution of the pure states, $\Sigma_r(s)$, is not a convex upward function in this system and does not agree with $\Sigma(s)$, which is evaluated by the relation of (21) using $g(x)$.

Next, we assessed the rate function for the region of $\alpha > \alpha_s$. In this region, the generation of samples that are perfectly compatible with the Ising perceptrons rarely occurs and is dominated by $s = 0$. Therefore, we numerically evaluated the probability that a given set of samples D^M could be separated by the Ising perceptron, P_{sep}, and estimated $R(0)$ as $R(0) = (1/N) \log P_{sep}$.

The resultant plots are given in figures 10 and 11 for $\alpha = 1.0$ and 1.5, respectively. The solid lines in these figures were obtained by the linear fitting for the numerical data. These figures show that the theoretical predictions are reasonably consistent with the values of extrapolation of the numerical data. The statistical errors are sufficiently small, and hence the differences between the analytical and numerical results should be the systematic errors due to the nonlinearity of the Ising perceptron.
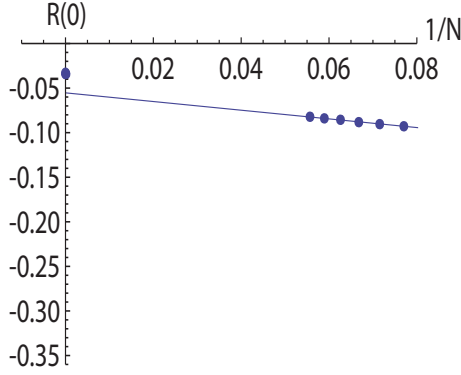
Figure 10. Size dependence of the rate function for $\alpha = 1.0$. The point at $1/N = 0$ is the value predicted by the frozen RSB solution. The system size increases from $N = 12$ to 18 in increments of 1. The data from $320,000$ samples were evaluated for each N. The statistical errors are smaller than the markers.
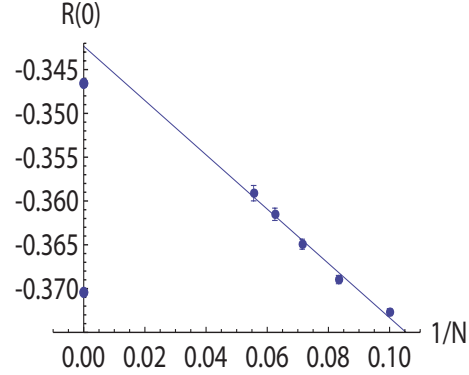
Figure 11. Size dependence of the rate function for $\alpha = 1.5$. The upper and lower plots at $1/N = 0$ are given by the RS2 and frozen RSB solutions, respectively. The system size increases from $N = 10$ to 18 in increments of 2. The data from $25,600,000$ samples were evaluated for each N.

6. Summary

In the present paper, we investigated the structure of the weight space of Ising perceptrons in which a set of random patterns is stored using the derivatives of the generating function of the partition function. This was achieved by carrying out a finite-n replica analysis under the assumption of one-step replica symmetry breaking (1RSB) handling Parisi's 1RSB parameter as a control parameter. For $\alpha < \alpha_{\mathrm{s}} = 0.833\ldots$, the analysis of $n \to 0$ indicates that the characteristic exponent of the size distribution of pure states is not convex upward, which implies that the weight space is equally dominated by a single large cluster of exponentially many weights and exponentially many clusters of a single weight. For $\alpha > \alpha_{\mathrm{s}}$, a set of random patterns is rarely compatible with the Ising perceptron. The $n \to 0$ analysis enables us to assess the rate function that characterizes a small probability that a cluster of a given entropy will emerge after the storage of random patterns. We found that a cluster of finite entropy is generated with a relatively high probability for $\alpha_{\mathrm{s}} < \alpha < \alpha_{\mathrm{GD}} = 1.245\ldots$, but this is very rare for $\alpha > \alpha_{\mathrm{GD}}$. These theoretical predictions have been validated by extensive numerical experiments. We also drew a complete phase diagram on the n-α plane, in which $(n, \alpha) = (1, \alpha_{\mathrm{GD}})$ becomes a tricritical point. The line $n = 1$ that passes through the tricritical point is analogous to the Nishimori line in the theory of spin glasses.

We stressed the use of the replica method as a tool for calculating the complexity and rate function. The developed formalism enables the extraction of useful information about typical and atypical behaviors of the objective system from a single generating function in an unified manner. It is hoped that the results of the present study will help

to clarify systems with complex phase spaces as well as the replica method itself.

Acknowledgments

Appendix A. Derivation of RS and 1RSB solutions

For $n = 1, 2, \ldots, \in \mathbb{N}$, the nth moment of Z is expressed as

$$[Z^n]_{D^M} = \sum_{\boldsymbol{S}^1, \boldsymbol{S}^2, \ldots, \boldsymbol{S}^n} \left[\prod_{a=1}^{n} \prod_{\mu=1}^{M} \Theta \left(-y_\mu \frac{\boldsymbol{S}^a \cdot \boldsymbol{x}_\mu}{\sqrt{N}} \right) \right]_{D^M}, \quad (A.1)$$

where the brackets $[\cdots]_{D^M}$ denote the average over the quenched randomness D^M. The variable $u_\mu^a = -y_\mu \boldsymbol{S}^a \cdot \boldsymbol{x}_\mu / \sqrt{N}$ ($a = 1, 2, \ldots, n; \mu = 1, 2, \ldots, M$) can be regarded as multivariate Gaussian random variable, which is characterized as

$$\left[u_\mu^a \right]_{D^M} = 0, \quad \left[u_\mu^a u_\nu^b \right]_{D^M} = \delta_{\mu\nu} \left(\delta_{ab} + (1 - \delta_{ab}) q^{ab} \right), \quad (A.2)$$

where $q^{ab} = (1/N) \sum_{i=1} S_i^a S_i^b$ ($a, b = 1, 2, \ldots, n$). This observation yields the following expression

$$[Z^n]_{D^M} = \sum_{\boldsymbol{S}^1, \boldsymbol{S}^2, \ldots, \boldsymbol{S}^n} \int \prod_{a<b} dq^{ab} \delta \left(\boldsymbol{S}^a \boldsymbol{S}^b - N q^{ab} \right) \left[\prod_{a=1}^{n} \prod_{\mu=1}^{M} \Theta \left(u_\mu^a \right) \right]_{D^M}$$

$$= \int \prod_{a<b} \frac{dq^{ab} d\widehat{q}^{ab}}{2\pi} \exp N \left(-\sum_{a<b} q^{ab} \widehat{q}^{ab} + \log \sum_{S^1, S^2, \ldots, S^n} e^{\sum_{a<b} \widehat{q}^{ab} S^a S^b} \right.$$

$$\left. + \alpha \log \left[\prod_{a=1}^{n} \Theta \left(u^a \right) \right]_{\boldsymbol{u}} \right), \quad (A.3)$$

where $[\cdots]_{\boldsymbol{u}}$ denotes the average with respect to the multivariate Gaussian variables the moments of which are given by (A.2). In order to derive the previous expression, we used the Fourier expression of the delta function

$$\delta \left(\boldsymbol{S}^a \boldsymbol{S}^b - N q^{ab} \right) = \int_{-i\infty}^{+i\infty} \frac{d\widehat{q}^{ab}}{2\pi} \exp \left(\widehat{q}^{ab} (\boldsymbol{S}^a \boldsymbol{S}^b - N q^{ab}) \right). \quad (A.4)$$

Applying the saddle-point method to (A.3), we immediately obtain (12). In order to investigate $\phi(n)$ for $n \in \mathbb{R}$, we need an ansatz on the form of the saddle point q^{ab}. We first adopt the RS ansatz

$$\widehat{q}^{ab} = \widehat{q}, \quad q^{ab} = q \ (a < b = 1, 2, \ldots, n). \quad (A.5)$$

Under this assumption, we obtain

$$\sum_{a<b} q^{ab}\widehat{q}^{ab} = \frac{1}{2}n(n-1)\widehat{q}q, \tag{A.6}$$

$$\sum_{S^1,S^2,\ldots,S^n} e^{\sum_{a<b}\widehat{q}^{ab}S^aS^b} = e^{-\frac{1}{2}n\widehat{q}}\int Dz(2\cosh\sqrt{\widehat{q}}z)^n. \tag{A.7}$$

Under the RS ansatz, the Gaussian variable u^a can be decomposed to two independent Gaussian variables of zero mean and unit variance x^a and z as

$$u^a = \sqrt{1-q}x^a + \sqrt{q}z. \tag{A.8}$$

Using this expression, we obtain

$$\left[\prod_{a=1}^{n}\Theta(u^a)\right]_{\boldsymbol{u}} = \int Dz\left(E\left(\sqrt{\frac{q}{1-q}}z\right)\right)^n, \tag{A.9}$$

where $E(u) = \int_u^{+\infty} Dz$. Using the above expressions, we obtain

$$\phi_{\mathrm{RS}}(n) = \underset{q,\widehat{q}}{\mathrm{Extr}}\left\{-\frac{1}{2}n(n-1)\widehat{q}q - \frac{1}{2}n\widehat{q} + \log\int Dz(2\cosh\sqrt{\widehat{q}}z)^n\right.$$
$$\left. + \alpha\log\int Dz\left(E\left(\sqrt{\frac{q}{1-q}}z\right)\right)^n\right\}, \tag{A.10}$$

The saddle point conditions are

$$q = \frac{\int Dz\cosh^n\sqrt{\widehat{q}}z\tanh^2\sqrt{\widehat{q}}z}{\cosh^n\sqrt{\widehat{q}}z}, \tag{A.11}$$

$$\widehat{q} = \frac{\alpha}{1-q}\frac{\int Dz\left(\frac{E'}{E}\right)^2 E^n}{\int DzE^n}, \tag{A.12}$$

where $E'(x) = dE(x)/dx = -e^{-x^2/2}/\sqrt{2\pi}$. Note that the arguments of E and E' are $\sqrt{q/(1-q)}z$. As previously noted, there are two solutions to (A.11) and (A.12), i.e., the **RS1** and **RS2** solutions presented in section 3.1.1.

Next, we use the 1RSB ansatz. The replica indices are divided into n/m groups of identical size m, and q^{ab} and $\widehat{q}^{ab}$ are parameterized as

$$(q^{ab},\widehat{q}^{ab}) = \begin{cases} (q_1,\widehat{q}_1) & (\ a \text{ and } b \text{ belong to the same group }) \\ (q_0,\widehat{q}_0) & (\text{ otherwise }) \end{cases} \tag{A.13}$$

This assumption yields

$$\sum_{a<b} q^{ab}\widehat{q}^{ab} = \frac{1}{2}n(m-1)\widehat{q}_1q_1 + \frac{1}{2}n(n-m)\widehat{q}_0q_0, \tag{A.14}$$

$$\sum_{S^1,S^2,\ldots,S^n} e^{\sum_{a<b}\widehat{q}^{ab}S^aS^b} = e^{-\frac{1}{2}n\widehat{q}_1}\int Dz_0\left(\int Dz_1(2\cosh h)^m\right)^{n/m}, \tag{A.15}$$

where $h = \sqrt{\widehat{q}_1-\widehat{q}_0}z_1 + \sqrt{\widehat{q}_0}z_0$. The Gaussian variable u^a can be decomposed to obtain

$$u^a = \sqrt{1-q_1}y_{\sigma a} + \sqrt{q_1-q_0}x_\sigma + \sqrt{q_0}z, \tag{A.16}$$

where $x_\sigma, y_{\sigma a}$ and z are independent Gaussian variables of zero mean and unit variance. The index σ indicates a block and a pair of σ and a specifies a replica in the σ block. This transformation yields the following expression:

$$\left[\prod_{a=1}^{n} \Theta\left(u^a\right)\right]_{\boldsymbol{u}} = \int Dz \left(\int Dx \left\{E(y_0(z,x))\right\}^m\right)^{n/m}, \tag{A.17}$$

where

$$y_0(z,x) = -\sqrt{\frac{q_0}{1-q_1}}z - \sqrt{\frac{q_1-q_0}{1-q_1}}x. \tag{A.18}$$

Finally, we obtain

$$\begin{aligned}
\phi_{\text{1RSB}}(n,m) = \operatorname*{Extr}_{q_1,q_0,\widehat{q}_1,\widehat{q}_0} &\left\{-\frac{1}{2}n(m-1)\widehat{q}_1 q_1 - \frac{1}{2}n(n-m)\widehat{q}_0 q_0 \right.\\
&-\frac{1}{2}n\widehat{q}_1 + \log\int Dz_0\left(\int Dz_1(2\cosh h)^m\right)^{n/m}\\
&\left. + \alpha\log\int Dz_0\left(\int Dz_1\left\{E(y_0(z_0,z_1))\right\}^m\right)^{n/m}\right\}. \tag{A.19}
\end{aligned}$$

Taking the extremization, we can derive the saddle-point equations. The possible solutions of the equations are discussed in section 3.1.2.

References

[1] Parisi G and Rizzo T 2008 *Phys. Rev. Lett.* **101** 117205
[2] Nakajima T and Hukushima K 2008 *J. Phys. Soc. Japan* **77** 074718
[3] Obuchi T, Kabashima Y and Nishimori H 2009 *J. Phys. A: Math. Gen.* **42** 075004
[4] Boffetta G, Monasson R and Zecchina R 1993 *J. Phys. A: Math. Gen.* **26** L507
[5] Monasson R and O'Kane 1994 *Europhys. Lett.* **27** 85
[6] Monasson R 1995 *Phys. Rev. Lett.* **75** 2847
[7] Derrida B 1981 *Phys. Rev.* B **24** 2613
[8] Gardner E and Derrida B 1989 *J. Phys. A: Math. Gen.* **22** 1975
[9] Nakajima T and Hukushima K 2009 *Phys. Rev.* E **80** 011103
[10] Monasson R and Zecchina R 1997 *Phys. Rev.* E **56** 1357
[11] Mézard M, Parisi G and Zecchina R 2002 *Science* **297** 812
[12] Montanari A, Ricci-Tersenghi F and Semerjian G 2008 *J. Stat. Mech.* P04004
[13] Zdeborová L and Krząkała F 2007 *Phys. Rev.* E 76 031131
[14] Gardner E and Derrida B 1988 *J. Phys. A: Math. Gen.* **21** 271
[15] Krauth W and Mézard M 1989 *J. Phys. France* **50** 3057
[16] Cocco S, Monasson R and Zecchina R 1996 *Phys. Rev.* E **54** 717
[17] Biroli G and Monasson R 2000 *Europhys. Lett.* **50** 155
[18] Ardelius J and Zdeborová L 2008 *Phys. Rev.* E **78** 040101
[19] Nishimori H 2001 *Statistical Physics of Spin Glasses and Information Processing: An Introduction* (Oxford: Oxford University Press)
[20] Mézard M and Montanari A 2009 *Information, Physics, and Computation* (Oxford: Oxford University Press)
[21] Bouchaud J P and Mézard M 1997 *J. Phys. A: Math. Gen.* **30** 7997
[22] Braunstein A and Zecchina R 2006 *Phys. Rev. Lett.* **96** 030201
[23] Horner H 1992 *Z. Phys. B* **86** 291

[24] Horner H 1992 *Z. Phys. B* **87** 371

[25] Györgyi G 1990 *Phys. Rev. A* **41** 7097

[26] Engel A and Weigt M 1996 *Phys. Rev. E* **53** 2064

[27] Nishimori H 1980 *J. Phys. C: Solid State Phys.* **13** 4071

[28] de Almeida J R L and Thouless D J 1978 *J. Phys. A: Math. Gen.* **11** 983

[29] Akaike H 1974 *IEEE Trans. on Automatic Control* **19** 716